Laplacian SVM

Máster Universitario en Ciencia de Datos - Métodos Funcionales en Aprendizaje Automático

Ángela Fernández Pascual

Escuela Politécnica Superior
Universidad Autónoma de Madrid

Academic Year 2021/22

Universidad Autónoma
de Madrid

Contents

Motivation

Motivation

Motivation
- To exploit the geometry of the data distribution.
- To incorporate this geometry as a regularization term.

Regularization terms
1. To penalize the classifier complexity in the *ambient space*.
2. To penalize the complexity measured by the distribution geometry.

Three concepts
1. Spectral graph theory - clustering & classification
2. Manifold Learning
3. Regularization in RKHS
   $\rightarrow$ kernel-based algorithms - classification & regression

$\Rightarrow$ Framework of methods from unsupervised to fully supervised.

- We focus on the semi-supervised case.

Meaning of Semi-supervised Learning

Meaning of Semi-supervised Learning

Semi-supervised Learning

It is an approach in between unsupervised and supervised learning.
It combines a small amount of **labeled** data with a large amount of **unlabeled** data during training.

Why using a semi-supervised method?

1. Practicality: easier to obtain unlabeled data.
2. It is a natural learning process.

Notebook

Semi-supervised Learning

Manifold Regularization

Manifold Regularization - Notation

Data $\mathscr{S} = \{\underbrace{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\ell)}}_{\text{labeled data}}, \underbrace{\mathbf{x}^{(\ell+1)}, \ldots, \mathbf{x}^{(N)}}_{\text{unlabeled data}}\} \qquad \mathbf{x}^{(i)} \in \mathbf{X} \subset \mathbb{R}^{M}.$

$$\mathscr{S} = \mathscr{L} \cup \mathscr{U} \text{ where } \begin{cases} \mathscr{L} = & \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}), i = 1, \ldots, \ell\} \\ \mathscr{U} = & \{\mathbf{x}^{(i)}, i = \ell + 1, \ldots, N\} \end{cases}.$$

Dimension $M = \ell + u.$

Labels $\mathbf{y}^{(i)} \in \{-1, 1\}.$

Laplacian $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}.$

Kernel $\mathbf{K} \in \mathbb{R}^{N,N},\ \mathcal{K}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$

Target function $f : X \to \mathbb{R}$

$$f = [f(\mathbf{x}^{(i)}), \mathbf{x}^{(i)} \in \mathscr{S}]^\top$$

Manifold Regularization (I)

- Following a **manifold learning** approach:
  - Two close points in the intrinsic geometry of the data distribution (geodesic distance on $\mathscr{M}$)

  $\rightarrow$ should have the same or similar label .
- Intrinsic regularizer $\|f\|_I^2$:

$$\|f\|_I^2 = \sum_{i=1}^{N} \sum_{j=i}^{N} w_{ij} \left(f(\mathbf{x}^{(i)}) - f(\mathbf{x}^{(j)}) \right)^2 = f^\top \mathbf{L} f.$$

- Regularization framework for **function learning**:
  - $\mathcal{K}(\cdot,\cdot)$ in a RKHS with norm $\|\cdot\|_A$.

Manifold Regularization (II)

- The optimal target function can be estimated as:

$$\boxed{f^* = \underset{f}{\operatorname{argmin}} \sum_{i=1}^{\ell} V(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}, f) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2}$$

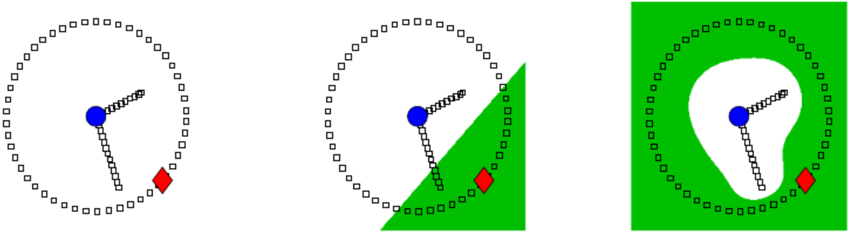
  - $V(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}, f) \rightarrow$ loss function.
  - $\gamma_A \|f\|_A^2 \rightarrow$ enforces smoothness on the solution (*ambient* norm).
  - $\gamma_I \|f\|_I^2 \rightarrow$ enforces smoothness on the sample $\mathscr{M}$ (*intrinsic* norm).

Manifold Regularization

$$f^* = \operatorname*{argmin}_f \sum_{i=1}^{\ell} V(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}, f) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2.$$

- f^* admits an expansion in terms of the N points of $\mathscr{S}$:

$$\boxed{f^*(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i^* \mathcal{K}(\mathbf{x}^{(i)}, \mathbf{x})}.$$



Laplacian Support Vector Machines

Laplacian Support Vector Machines (I)

- In a traditional SVM:
  - $V(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}, f) \equiv \mathrm{L}_1$ loss
  - The well classified labels are not penalized.
- Minimization problem:

$$f^* = \operatorname*{argmin}_f \sum_{i=1}^{\ell} \boxed{\max(1 - \mathbf{y}^{(i)} f(\mathbf{x}^{(i)}), 0)} + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2$$

- f^* also admits an expansion:

$$f^*(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i^* \mathcal{K}(\mathbf{x}^{(i)}, \mathbf{x}) + \boxed{\mathbf{b}}.$$

Laplacian Support Vector Machines (II)

$$f^* = \operatorname*{argmin}_f \sum_{i=1}^{\ell} \max(1 - \mathbf{y}^{(i)} f(\mathbf{x}^{(i)}), 0) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2$$

- Constraint problem by introducing the slack variable ξ_i, and the expansion of f:

$$\begin{aligned} &\operatorname*{argmin}_{\alpha,\xi} \sum_{i=1}^{\ell} \xi_i + \gamma_A \alpha^\top \mathbf{K}\alpha + \gamma_I \alpha^\top \mathbf{KLK}\alpha \\ &s.t. \quad \mathbf{y}^{(i)}\Big(\sum_{j=1}^{N} \alpha_i \mathcal{K}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + b\Big) \geq 1 - \xi_i \qquad i = 1, \cdots, \ell \\ &\qquad\ \ \xi_i \geq 0 \qquad i = 1, \cdots, \ell \end{aligned}$$

Laplacian Support Vector Machines (III)

- Lagrange multipliers β, ς:

$$
\begin{aligned}
L_g(\alpha, \xi, b, \beta, \varsigma) = \sum_{i=1}^{\ell} \xi_i + \frac{1}{2}\alpha^\top (2\gamma_A \mathbf{K} + 2\gamma_I \mathbf{KLK})\alpha \\
- \sum_{i=1}^{\ell} \beta_i (\mathbf{y}^{(i)} (\sum_{j=1}^{N} \alpha_i \mathcal{K}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) + b) - 1 + \xi_i) \\
- \sum_{i=1}^{\ell} \varsigma_i \xi_i .
\end{aligned}
$$

- $\frac{\partial L_g}{\partial b} = 0 \Rightarrow \sum_{i=1}^{\ell} \beta_i \mathbf{y}^{(i)} = 0,$
- $\frac{\partial L_g}{\partial \xi_i} = 0 \Rightarrow 1 - \beta_i - \varsigma_i = 0 \Rightarrow 0 \leq \beta_i \leq 1.$

Laplacian Support Vector Machines (IV)

- For rewriting the Lagrangian as a function of α and β, we define:
  - $\mathbf{J}_{\mathscr{L}} = [\mathbf{I}\ \mathbf{0}] \in \mathbb{R}^{\ell,N}$, where $\mathbf{I} \in \mathbb{R}^{\ell,\ell}$; $\mathbf{0} \in \mathbb{R}^{\ell,u}$.
  - $\mathbf{Y} \in \mathbb{R}^{\ell,\ell}$ diagonal (labels).
- New Lagrangian:

$$
\begin{aligned}
L_g(\alpha,\beta) =& \frac{1}{2}\alpha^\top(2\gamma_A\mathbf{K} + 2\gamma_I\mathbf{KLK})\alpha - \sum_{i=1}^{\ell}\beta_i(\mathbf{y}^{(i)}(\sum_{j=1}^{N}\alpha_i\mathcal{K}(\mathbf{x}^{(i)},\mathbf{x}^{(j)}) + b) - 1) \\
=& \frac{1}{2}\alpha^\top(2\gamma_A\mathbf{K} + 2\gamma_I\mathbf{KLK})\alpha - \alpha^\top\mathbf{K}\mathbf{J}_{\mathscr{L}}^\top\mathbf{Y}\beta + \sum_{i=1}^{\ell}\beta_i.
\end{aligned}
$$

- Derivative with respect to α:

$$
\begin{aligned}
\frac{\partial L_g}{\partial\alpha} = 0 \Rightarrow & (2\gamma_A\mathbf{K} + 2\gamma_I\mathbf{KLK})\alpha - \mathbf{K}\mathbf{J}_{\mathscr{L}}^\top\mathbf{Y}\beta = 0 \\
\Rightarrow & \alpha = (2\gamma_A\mathbf{I} + 2\gamma_I\mathbf{KL})^{-1}\mathbf{J}_{\mathscr{L}}^\top\mathbf{Y}\beta.
\end{aligned}
$$

Laplacian Support Vector Machines (V)

- Dual problem:

$$\beta^* = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{\ell} \beta_i - \frac{1}{2}\beta^\top \mathbf{Q}\beta$$

$$s.t. \sum_{i=1}^{\ell} \beta_i \mathbf{y}^{(i)} = 0$$

$$0 \leq \beta_i \leq 0 \qquad i = 1, \cdots, \ell$$

where

$$\mathbf{Q} = \mathbf{Y}\mathbf{J}_{\mathscr{L}}\mathbf{K}(2\gamma_A \mathbf{I} + 2\gamma_I \mathbf{K}\mathbf{L})^{-1}\mathbf{J}_{\mathscr{L}}^\top \mathbf{Y}.$$

- With this expression we can train the LapSVM and obtain the expansion for f^*.

Notebook

LapSVM

Conclusions

- RKHS, manifold learning and spectral methods have been mixed.
- The geometric structure of the sample data has been incorporated into the regularization framework.

$\Rightarrow$ Definition of a semi-supervised model (LapSVM).

References

References

- Belkin, M., Niyogi, P., Sindhwani, V. (2006). **Manifold regularization: A geometric framework for learning from labeled and unlabeled examples**. Journal of machine learning research, 7(11).
- Melacci, S., Belkin, M. (2011). **Laplacian Support Vector Machines Trained in the Primal**. Journal of Machine Learning Research, 12(3).